

Molecular dynamics simulations of ASD-associated SHANK3 variants

Stephan Niebling* MB Michael Bucher MM Marina Mikhaylova

Updated date: Dec 17, 2021

*For correspondence: stephan.niebling@embl-hamburg.de

An abbreviated version of this protocol was published in eLIFE in May 2021

Autism-associated *SHANK3* missense point mutations impact conformational fluctuations and protein turnover at synapses

DOI: 10.7554/eLife.66165

Detailed protocol

Required software and sources

All required software is freely available:

- Preparation of mutant structures (R12C and L68P): UCSF Chimera (Pettersen et al., 2004)
- MD simulations using the Amber03 force-field: GROMACS (Duan et al., 2003; Hess et al., 2008; Pronk et al., 2013)
 - Gromacs is freely available here: gromacs.org
 - Recommended online tutorial: <http://www.mdtutorials.com/gmx/>
- For data analysis and plots: Python and its modules numpy (Harris et al. 2020), scipy (Virtanen et al. 2020), matplotlib (Hunter 2007).
- Visualization of structures and data analysis: VMD 1.9.3. (Humphrey et al., 1996)

Included files

- MD parameter (mdp) files: these files are needed to prepare the run input files for each MD step using "gmx grompp" (adapted from <http://www.mdtutorials.com/gmx/>)
- Starting structures:
 - Wild-type structure 5G4X and corresponding protein-only structure (without waters and additives)
 - Mutant structures: L68P and R12C (created with Chimera)
- TCL script for VMD used for RMSF calculations:
 - See also <https://www.ks.uiuc.edu/Research/vmdl/>

Preparation of starting structures

As starting structure for the wild-type system we have used the pdb entry 5G4X from <https://www.rcsb.org/>

The mutants R12C and L68P were generated by replacing the respective residues in UCSF Chimera using the Dunbrack 2010 backbone-dependent rotamer library (Shapovalov & Dunbrack, 2011). Alternatively, other programs such as Pymol can be used for this step. Individual side chains can be replaced in Chimera using the Rotamers tool (Tools... Structure Editing). Thereby individual side chain rotamers can be selected based on their predicted probabilities and can be incorporated into the corresponding protein structure. This, however, does not modify the backbone and may initially produce clashes or steric repulsions such that the generated mutants require subsequent minimization.

Preparation of MD systems

The following instructions are for the wild type (WT) system. The mutant systems can be prepared using the same protocol. The only difference is the charge of the protein system, i.e. the number of ions that are required to neutralize the system.

Preparation of the MD simulation box

- Create a folder named WT and copy the mdp input files (folder in attachment) and 5G4X.pdb (from protein database) into this folder
- Clean pdb by removing all lines starting with HETATM from 5G4X.pdb using a text editor: name the new pdb 5G4X_onlyprot.pdb
- Create Gromacs coordinate file prot.gro from pdb
 - `gmx pdb2gmx -f 5G4X_onlyprot.pdb -o prot.gro`
 - Choose Amber03 as force field (here number 1) and TIP3P as explicit water model (here number 1)
 - Total charge -1 (for R12C this will be -2)

- Note: Depending on the protonation criteria of the program used, different charges may be obtained.
- Place structure in box with 1 nm distance of the protein to each box face
 - `gmx editconf -f prot.gro -o box.gro -c -d 1 -bt cubic`
- Fill box with water molecules
 - `gmx solvate -cp box.gro -cs spc216.gro -p topol.top -o solv.gro`
- Add ions to neutralize charge: In this case we need to add 1 Na⁺ ion
 - First create the input file with grompp. There will be one warning about the net charge being not zero. We can ignore this one since we will add counterions in the second step
 - `gmx grompp -f ions.mdp -c solv.gro -p topol.top -o ions.tpr -maxwarn 1`
 - Now run the actual ion placement job choosing the "SOL" group as solvent atoms (here number 13)
 - `gmx genion -s ions.tpr -o ions.gro -p topol.top -pname NA -nname CL -np 1`

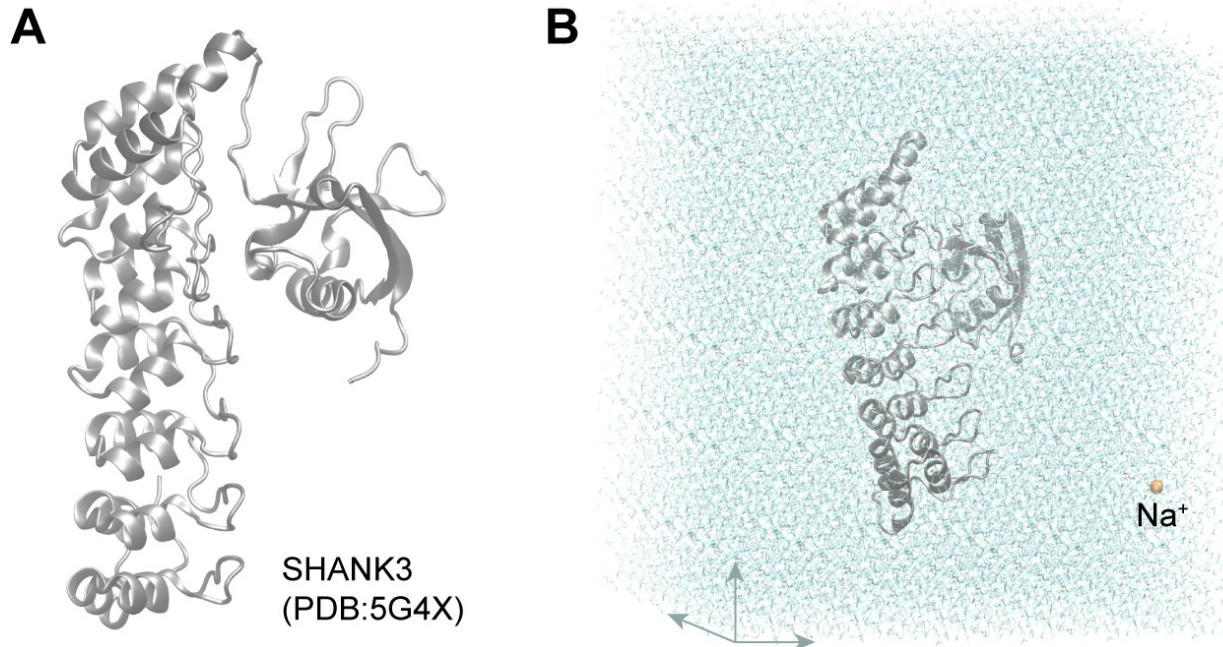


Fig 1: Solvation and charge neutralization of an N-terminal fragment of SHANK3

(A) The X-ray structure of the N-terminal fragment of SHANK3 including the SPN- and ARR-domain is shown (PDB: 5G4X; Lilja et al., 2017). Both domains are connected by a flexible linker. (B) After preparation of the MD system, the protein is solvated in a cubic box and charge neutralized by the addition of a Na⁺ ion. Images were created using VMD 1.9.3.

Equilibration of the MD simulation box

- Set up and run minimization. The output files will have the prefix em. One may (optionally, requires CUDA) use the GPU for nonbonded interactions with the flag "-nb gpu", specify the number of threads used with the -nt flag and show more information during the run with -v. In case you run your simulation on a node/computer that is shared, the -pin on option avoids pinning threads from different mdrun instances to the same core. These optional parameters apply to all following mdrun commands.
 - `gmx grompp -f minim.mdp -c ions.gro -p topol.top -o em.tpr`
 - `gmx mdrun -v -nb gpu -nt 8 -pin on -deffnm em`
- Prepare and run 100 ps of NVT equilibration
 - `gmx grompp -f nvt.mdp -c em.gro -p topol.top -o nvt.tpr -r em.gro`
 - `gmx mdrun -deffnm nvt`
- Prepare and run 100 ps of NPT equilibration
 - `gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p topol.top -o npt.tpr -r nvt.gro`
 - `gmx mdrun -deffnm npt`
- Now your system is ready for the production run

Preparation of production run

- Set up and run a 1000 ns MD simulation (this was done on a high performance computer)
 - `gmx grompp -f md_long.mdp -c npt.gro -t npt.cpt -p topol.top -o md_long.tpr`
 - `gmx mdrun -deffnm md_long -v -pin on -nt 8 -nb gpu`
- Create separate folders for the mutants R12C and L68P and repeat all the aforementioned steps.
 - Note: R12C will have a different charge which has to be taken into account in the neutralization step (cf. Preparation of the MD simulation box)

Processing of MD simulation prior to analysis

Before analyzing the trajectory it is necessary to center the protein molecule in the box (remove jumps in the structure from one side of the box to the other) and to remove rotation and translation (optional) in order to visually inspect the trajectory in VMD. We recommend this prior to analyzing the RMSF and RMSD shown in the next section.

- Create a trajectory with the protein centered (based on the C-alpha atoms). For our analysis we only need the protein coordinates, which saves space. Therefore we run the following command with the C-alpha group (typically 3) for centering and the protein group (typically 1) as output

- `gmx trjconv -pbc mol -center -s md_long.tpr -f md_long.xtc -o md_long_cent.xtc`
- Create a centered structure that can be used later with VMD. Again we choose the C-alpha group (3) for centering and the protein group (1) for output using the following command:
 - `gmx trjconv -f md_long.tpr -pbc mol -center -o md_long_rottrans.gro -s md_long.tpr`
- The final step of processing the trajectory is to remove rotation and translation of the protein. For the least squares fit we use the C-alpha group (3) and for output the protein group (1):
 - `gmx trjconv -s md_long.tpr -f md_long_cent.xtc -o md_long_rottrans.xtc -fit rot+trans`

Analysis of MD simulations

RMSF analysis

RMSF analysis can be performed in multiple ways either using gromacs or an external software such as VMD 1.9.3. Subsequently, RMSF analysis with VMD is illustrated. First load the centered trajectory without rotation and translation into VMD using “prot.gro” to define the molecule and “md_long_rottrans.xtc” to define the trajectory. For RMSF and RMSD analysis, every fifth frame was loaded (Stride: 5). To perform RMSF analysis, copy the file “RMSF.TCL” (provided as a supplementary file) to the VMD installation directory. By default, this is /usr/local/lib/vmd on Unix systems, and C:\Program Files\University of Illinois\VMD on Windows systems. Within this file, adjust the frame numbers in the actual RMSF measurement command depending on your total number of frames:

- `set rmsf [measure rmsf $sel first 0 last 20002 step 1]`

Save the adjusted file and run the script from the Tk console within VMD using the command:

- `source rmsf.tcl`

A text file named “RMSF.txt” will be saved in the same working directory containing two columns. The first column represents the protein residue number and the second column gives the measured RMSF values in Ångström. The data can subsequently be plotted with any graphing software.

Alternative RMSF calculation using gromacs tools

To perform RMSF analysis with gromacs first create an index file:

- `gmx make_ndx -f md_long_rottrans`

We will then use the trajectory to calculate the RMSF for the protein using the C-alpha group (usually selection 3). The -res flag will result in an output of RMSF vs. residue number (instead of atom number):

- `gmx rmsf -f md_long_rottrans.xtc -s md_long_rottrans.gro -n index.ndx -res`

The created output file rmsf.xvg contains two columns: the first one are the residue numbers, the second one are the RMSF values. This data can be loaded and visualized with a program of your choice. One quick way to visualize it is with xmgrace (needs to be installed):

- `xmgrace rmsf.xvg`

RMSD analysis

Within VMD, RMSD analysis can be performed using the RMSD Trajectory Tool (Extensions > Analysis > RMSD Trajectory Tool). Therefore, the same trajectory is used as for the RMSF analysis (md_long_rottrans.xtc; Stride: 5). Within the RMSD Trajectory Tool make sure to define the selected atoms as “protein”. To analyze the RMSD only for a part of the protein (e.g. an individual domain), specify the corresponding residues instead (e.g. “residue 1 to 92”). Then specify the following parameters:

- Selection modifiers = backbone
- Reference mol = Top
- Trajectory ref frame = 1

Subsequently the measured RMSD values can be saved as separate file as plotted with any graphing software.

Alternative RMSD calculation using gromacs tools

For obtaining the RMSD using the gromacs tool rms use the following command:

- `gmx rms -s md_long.tpr -f md_long_rottrans.xtc -o rmsd.xvg -tu ns`
- Choose C_alpha both for the least-squares fit and the subsequent RMSD calculation

The created output file rmsd.xvg contains two columns: the first one contains the simulation time in ns, the second one are the RMSD values (in nm). Similarly to the RMSF values, the RMSD can for instance be visualized using xmgrace:

- `xmgrace rmsd.xvg`







References

- Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., & Kollman, P. (2003). A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *Journal of Computational Chemistry*, 24(16), 1999–2012. <https://doi.org/10.1002/jcc.10349>
- Harris, C. R. et al. Array programming with NumPy. *Nature* 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2> (2020).
- Hess, B., Kutzner, C., Van Der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3), 435–447. <https://doi.org/10.1021/ct700301q>
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14(October 1995), 33–38.

<https://www.tapbiosystems.com/tap/products/index.htm>

- Hunter, J. D. Matplotlib: A 2d graphics environment. Comput. Sci. Eng. 9, 90–95. <https://doi.org/10.1109/mcse.2007.55> (2007).
- Lilja, J., Zacharchenko, T., Georgiadou, M., Jacquemet, G., De Franceschi, N., Peuhu, E., Hamidi, H., Pouwels, J., Martens, V., Nia, F. H., Beifuss, M., Boeckers, T., Kreienkamp, H. J., Barsukov, I. L., & Ivaska, J. (2017). SHANK proteins limit integrin activation by directly interacting with Rap1 and R-Ras. Nature Cell Biology, 19(4), 292–305. <https://doi.org/10.1038/ncb3487>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shiels, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics, 29(7), 845–854. <https://doi.org/10.1093/bioinformatics/btt055>
- Shapovalov, M. V., & Dunbrack, R. L. (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure, 19(6), 844–858. <https://doi.org/10.1016/j.str.2011.03.019>
- Virtanen, P. et al. SciPy 1.0: Fundamental algorithms for scientific computing in python. Nat. Methods 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2> (2020).

Related files

 mdp_files_gromacs.zip	
 Scripts.zip	
 starting_structures.zip	

How to cite: (Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Niebling, S., Bucher, M. and Mikhaylova, M. (2021). Molecular dynamics simulations of ASD-associated SHANK3 variants. Bio-protocol Preprint. bio-protocol.org/prep1479.
2. Bucher, M., Niebling, S., Han, Y., Molodenskiy, D., Hassani Nia, F., Kreienkamp, H., Svergun, D., Kim, E., Kostyukova, A. S., Kreutz, M. R. and Mikhaylova, M. (2021). Autism-associated *SHANK3* missense point mutations impact conformational fluctuations and protein turnover at synapses. eLIFE. DOI: [10.7554/eLife.66165](https://doi.org/10.7554/eLife.66165)

Copyright: Content may be subjected to copyright.